

Biospheric & Ecological Forecasting

Case Study: Wildfire Prediction

David Danks

NASA IDU/AR Workshop
February 5, 2004



Project Motivation



- General goals:
 - A (relatively) domain-general system for ecological / biospheric prediction and forecasting
 - Automated causal discovery for skewed data
- Reasons to focus on wildfire:
 - Instance of an important class of problems: Predicting rare events in very large datasets
 - Significant economic, ecological, and aesthetic costs (suppression alone costs > \$1B/year)

Difficulty of Wildfire Prediction

- Measurement problems:
 - Many known causes are unmeasured and essentially random
 - E.g., arson, carelessness, etc.
 - Other (potentially) causally relevant variables are measured, but are not easily accessible
 - E.g., meteorological measurements, local vegetation types, elevations/slopes, etc.
-

Difficulty of Wildfire Prediction

- Statistical problems:
 - Wildfire is *extremely* rare (despite the press)
 - <Weather – ecology – fire> joint distribution is non-stationary
 - Retrodiction and explanation are much easier than prediction
 - Time series are *highly* non-Markov
 - Prediction at time t requires measurements from (at least) $t - 1$ day, ..., $t - 1$ year
-

Wildfire Prediction Problems

Given ecological and meteorological data about a particular spatiotemporal region, can we improve fire occurrence predictions?

- U.S. Forest Service has fire risk indices, but there has been no systematic study of their reliability
- Highly accurate predictions are an unrealistic goal
Even 5-10% improvement would be worthwhile

(Note: The primary focus of previous wildfire research has been predicting how an existing fire will spread)

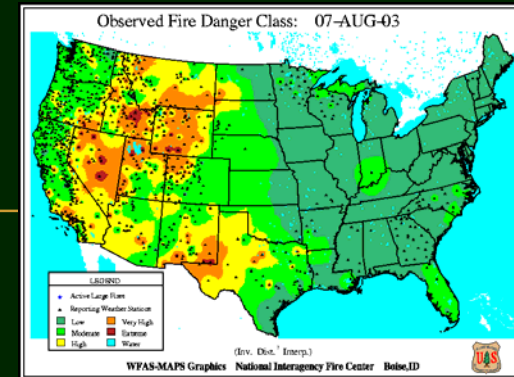
Three Lines of Attack

1. Using USFS data

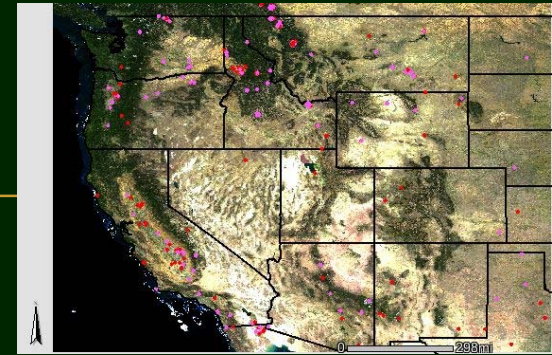
- ❑ Relevant variables are differently gridded and distributed over several large databases
- ❑ Analysis software designed for local managers
- ❑ Mixture of ground-based meteorological and local ecological variables

2. Using remote sensing (MODIS) data

3. Predicting frequency-area distributions

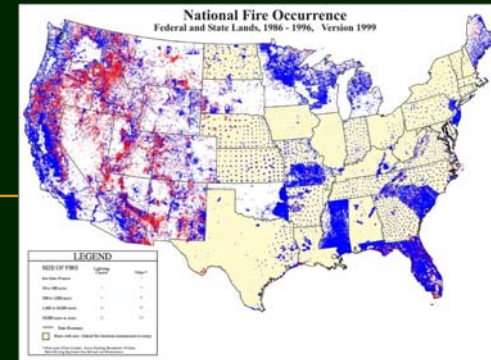


Three Lines of Attack



1. Using USFS data
2. Using remote sensing (MODIS) data
 - ❑ Global coverage on 1 km² scale every 1-2 days
 - ❑ Fire product known to be conservative (low false positive, but high false negative)
 - ❑ Predict the fire product using other products
3. Predicting frequency-area distributions

Three Lines of Attack



1. Using USFS data
2. Using remote sensing (MODIS) data
3. Predicting frequency-area distributions
 - ❑ National database of fires from 1986-1996
 - ❑ Predict the causes of regional fire occurrence variation at a coarser resolution
 - ❑ I.e., find causes of stable power law parameters for (regional) frequency-area distributions

Prediction Using USFS Data

- Two types of predictor variables:
 1. Meteorological / Ecological measurements (including indirect location and season data)
 2. Indices from Nat'l. Fire Danger Rating System (major decision tool for local fire managers)
 - Subsets of fires for prediction:
 1. All fires and all naturally occurring fires
 2. Large fires and large, naturally occurring fires
 - Large is >100 acres; 3% of fires, but 97% of burn area
-

Prediction Using USFS Data

- For all combinations of variables/subsets:
 - Matched sample – for each reported fire, match a non-fire for the same area in a previous year
 - Separate the data into training/test on 1/1/1997
 - Try different machine learning techniques:
 1. Logistic regression
(used by Andrews, *et al.*, but not by fire managers)
 2. Neural nets with linear/sigmoidal outputs
 3. Neural nets with RBF hidden units
 4. Binary classification trees
-

Results Using USFS Data

Do other machine learning methods predict fire occurrence better than logistic regression?

Not really.

In all conditions, regression was comparable to the non-regression technique(s) that performed best on validation data

Results Using USFS Data

Do the NFDRS fire risk indices help predict future fire occurrence?

Results Using USFS Data

Do the NFDRS fire risk indices help predict future fire occurrence?

No.

Average % correct classification (fire/non-fire):

Fire set \Rightarrow	All	Natural	Large	Lg. Natural
Input \Downarrow				
Weather	59.0	61.5	66.4	63.8
Risk Indices	54.2	54.9	61.0	60.3
Both	59.0	61.6	67.4	65.9

Prediction Using MODIS Data

- Predict current 8-day Fire product with:
 1. Preceding 8-day Leaf Area Index (LAI) Product
 2. Weighted sum of 44 previous 8-day LAI
 3. Separate measurements of previous 8-day LAI
 - Use several machine learning techniques:
 1. Logistic Regression
 2. Decision Trees
 3. Generalized Additive Model (GAM)
-

Results Using MODIS Data

- Weighted sum of previous LAIs is a much better predictor than single previous LAI
 - In fact, the weights are almost equal – strong evidence of highly non-Markovian data
 - Prediction in a particular 8-day period is quite good (% correct \approx 83%) using GAMs
 - Predictive power of models is currently unknown
-

Predicting Freq.-Area Distributions

- Model the fire frequency vs. fire size distribution as a power law:

$$Fr\left(\frac{\text{Fire of size } A}{\text{Time period } \Delta t}\right) = \beta \times A^{-\alpha}$$

- For a given spatiotemporal region,
 1. Is this power law a good model?
 2. What are the meteorological and ecological causes of α and β in that region?

Results for Freq.-Area Distributions

- Fit model to yearly fire occurrence for tiled Western U.S. and merge “similar” neighbors:
 - 188 $1^\circ \times 1^\circ$ squares with enough fires to model
 - ~23 coherent regions (~9 regions > 1 square)
 - For whole Western U.S., weighted $R^2 \approx 0.73$
 - *But* substantial parameter variation both (i) across years within a region; and (ii) across regions
 - Question: What are the meteorological / ecological causes of this variation?
-

Conclusions

- Predicting fire occurrence is a *hard* problem
 - ▣ Noisy and coarse measurements of highly stochastic variables – if they even exist!
 - But not *too* hard:
 - ▣ We can predict significantly better than chance
 - ▣ We can predict better than the current primary decision tool of the USFS
 - ▣ We have prospects for understanding causal mechanisms influencing fire size frequencies, and using that understanding for forecasting
-

Some Future Work

■ Extensions:

- Automated prediction on a national scale of fire occurrence using USFS & TOPS data
- Expanded MODIS analysis for prediction of fire occurrence using remote sensing data
- Application of causal discovery algorithms to find causes of parameters of frequency-area distributions

■ Novel problems/challenges:

- Prediction of fire risk indices at unmeasured sites using remote sensing
 - Estimating forest treatments from remote sensing data (and then using those estimated values in predictive models)
 - Application of lessons learned here to novel ecological / biospheric forecasting problems
-

Project Members

- Institute for Human & Machine Cognition (UWF)
 - Brian Bonnlander
 - Tianjiao Chu
 - David Danks (also Carnegie Mellon)
 - Clark Glymour (also Carnegie Mellon)
 - Fred Seibel
 - Frank Wimberly
 - NASA Ames collaborators
 - Keith Golden, Forrest Melton, Rama Nemani, Petr Votava
-

Acknowledgements

- Domain-specific support provided by:
 - Craig Allen & Cay Beely (USGS Jemez Mtn. Field Station)
 - Randy Balice & Brad McKown (LANL Office of Environmental Information Management)
 - Erik Martinson (Western Forest Fire Res. Center at CSU)
 - Mark Finney & Kristen Schmidt (USDA Forest Service, Rocky Mountain Research Station)
 - Jacqueline Klaver (USFS EROS Data Center)
 - Patrick McCarthy (The Nature Conservancy)
-